ANALYSIS OF UNCERTAINTY IN MODEL PREDICTIONS FOR LAKE LANIER

O.O. Osidele[1] and M.B. Beck[2]

AUTHORS: [1]Graduate Student and [2]Professor and Eminent Scholar, Environmental Informatics and Control Program, Warnell School of Forest Resources, The University of Georgia, Athens, Georgia 30602.
REFERENCE: Proceedings of the 2001 Georgia Water Resources Conference, held March 26-27, 2001, at the University of Georgia. Kathryn J. Hatcher, editor, Institute of Ecology, The University of Georgia, Athens, Georgia.

Abstract. In this paper we summarize a *comprehensive* analysis of model-based predictions that accounts for uncertainties due to: (i) the internal process parameters, (ii) the temporal pattern of variation of external inputs and forcing functions, and (iii) the initial state of the system. We apply the Regionalized Sensitivity Analysis (RSA) and Tree-Structured Density Estimation (TSDE) procedures to a generalized ecosystem model of Lake Lanier, Georgia. Our results suggest that the seasonal variation in watershed inputs, water temperature, and the timing of life-history stages of top predators, are as critical to predicting lake eutrophication as the internal ecosystem processes. These procedures are useful in ranking priorities for policy actions and directing the focus of scientific research on environmental systems.

INTRODUCTION

In a previous study, we employed the Regionalized Sensitivity Analysis procedure for identifying the key model parameters that match specified behaviors for Lake Oglethorpe, Georgia (Osidele and Beck, 1999). Although the results indicate apparent structural changes between observed past and speculated future behaviors, the analysis did not account for the effects of changes in the future pattern of external inputs and forcing functions. This oversight is quite common with model predictions that inform policy actions. Since model calibration is essentially parameter estimation, it seems reasonable to analyze prediction uncertainty also in terms of the parameters. However, if, for example, eutrophication control focuses primarily on regulating *external* nutrient loading, it seems hardly appropriate to analyze uncertainty of the informing model predictions in terms of parameters that characterize *internal* process mechanisms. Parameter uncertainty is just one of three major sources of prediction uncertainty (Beck, 1991) – the other two being, the assumed pattern of future inputs, and the estimated initial condition of the system at the start of the prediction.

The ecology of lakes and reservoirs is considerably influenced by external phenomena such as watershed hydrology, nutrient and sediment loading, atmospheric deposition and wind action (among many others), each of which is subject to future change, and therefore must be predicted as well. Thus, the future pattern of inputs is likewise uncertain. The initial state of the system is set either from direct measurement (subject to errors in instruments and field procedures), or estimated from prior models (likewise subject to uncertainty).

Our aim therefore is to "parameterize" these three sources of model uncertainty and combine them within a comprehensive analysis that evaluates the relative contribution of each to predicting eutrophication in Lake Lanier. To minimize the ambiguity of strictly predicting an unknown future, and to demonstrate the practical utility of our methods, we adopt the approach of specifying a single predictive task for the model. Our task specification for this case study is defned in terms of three commonly observed and frequently regulated indicators of lake eutrophication in the Southeastern U.S. Piedmont: phytoplankton, soluble phosphorus, and organic carbon concentrations.

METHODS

The Model

Figure 1 gives the conceptual structure of a dynamic ecosystem model of Lake Lanier, which we built using the MATLAB/Simulink software package. Each of the 13 state variables represents a functional group in a typical aquatic food web. In order to account for vertical gradients in soluble phosphorus during summer stratification, we apply the "simplest seasonal approach" (Simons and Lam, 1980) that separates the water column into two vertical compartments: the (top) epilimnion with soluble phosphorus and phytoplankton, and the (bottom) hypolimnion with soluble phosphorus, microbes and detritus.

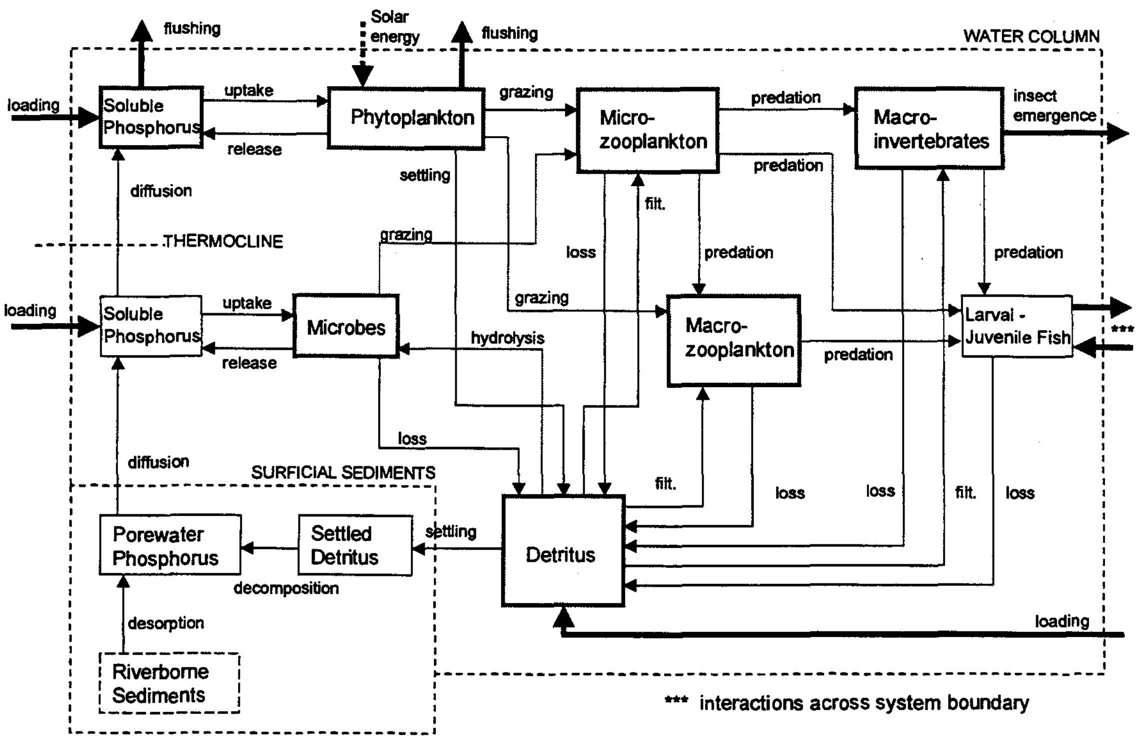

Figure 1. Conceptual structure of a generalized ecosystem model for Lake Lanier

The remaining state variables are modeled for the entire water column. Suspended solids (omitted from Figure 1 for clarity) contribute to light extinction in the water column, and settle into the sediment layer. Due to the daily time scale of our model, planktivorous larval-juvenile fish interact with a fixed adult fish biomass through piscivory and recruitment.

The model comprises 65 internal process parameters that include rate-constants (growth, feeding, mortality, metabolic losses, insect emergence, fish recruitment and detritus decomposition), half-saturation constants, assimilation efficiencies, light extinction coefficients and temperature constants. Other process parameters are settling velocities, diffusion coefficients and a rate constant for desorption of sediment-bound phosphorus.

The external inputs and forcing functions include tributary streamflow (loading phosphorus, suspended solids and detritus), incident solar radiation, and water temperature. For the purpose of our analysis these functions are parameterized as follows:

- streamflow: 35 different annual inflow patterns are extracted from the 43 years of data (1958-2000) available on the Corps of Engineers website;
- solar radiation: likewise, an index is assigned to the four years (1996-99) of daily data for Gainesville;
- water temperature: a sine function is fitted to data from the Clean Lakes Program (Hatcher, 1994) in terms of three parameters – the annual mean, range, and offset from the first day;
- in addition, insect emergence and fish spawning are parameterized by time of onset and duration.

Regionalized Sensitivity Analysis (RSA)

The RSA (Spear and Hornberger, 1980) is a Monte Carlo based method that starts with a random sampling of numerous parameter sets from a predefined range of values. Each set is fed into the model to produce an output, which is then compared to the real system using a set of goodness-of-fit conditions (the *behavior definition*). A behavior definition can be any of the following: (i) an interpretation of past observations, (ii) speculation about future behavior, (iii) existing or proposed regulatory standards, or (iv) an expression of stakeholder concerns and desires for the condition of the system. For this study the behavior definition – the task specification for model prediction – covers the May-September period of the annual cycle. In Lake Lanier, this most critical period for water quality coincides with summer stratification, peak primary production, hypolimnetic dissolved oxygen depletion, fish spawning, and the peak recreational season. Water quality behavior is defined as follows:

- soluble phosphorus: peak concentration should not exceed $7\mu g.L^{-1}$ in the epilimnion and $16\ \mu g.L^{-1}$ in the hypolimnion; mean epilimnion concentration should not exceed 55% of the hypolimnion mean;
- phytoplankton: peak and mean biomass (expressed as carbon) should not exceed $2.0mg.L^{-1}$, and $1.3mg.L^{-1}$, respectively; and
- total organic carbon: peak concentration should not exceed $2.4\ mg.L^{-1}$.

Model outputs that match the behavior definition (*B*) are distinguished from those that do not (*NB*), and the respective parameter values are subjected to a test of statistical difference. The importance (sensitivity) of a model parameter is directly related to the significance of the difference between the marginal distributions of its *B* and *NB* values. The model parameters are grouped into three sensitivity classes – (*1*) critical, (*2*) important and (*3*) insignificant.

Tree-Structured Density Estimation (TSDE)

The TSDE (Spear et al., 1994) addresses a critical weakness of the RSA procedure – that it uses *marginal* parameter distributions, which discount the correlations among parameter values. While a significant difference between the *B* and *NB* values is sufficient to indicate parameter sensitivity, the converse is not always true, because each correlating parameter exhibits a flattened marginal distribution which does not clearly distinguish between the *B* and *NB* values.

The TSDE performs a *multivariate* analysis of the *B* parameter values that indicates the relative importance of individual model parameters. Using simple density estimation, the RSA sampling domain is recursively partitioned into low- and high-density subdomains. This process is based on the principle of constructing a one-dimensional histogram. Each subdomain is constructed in a *binary* fashion by splitting its parent domain on the axis of the parameter that produces the largest increase in overall accuracy of the density estimate for the multivariate distribution. This produces a binary tree structure, where the nodes represent the subdomains and the branches (the splits) are determined by the most sensitive model parameters.

Tracing a high-density terminal node from the root node is equivalent to locating those regions of the parameter space most likely to produce model outputs that match the behavior definition. Thus, the number of high-density terminal nodes that each parameter helps define is indicative of its relative importance.

RESULTS & DISCUSSION

Table 1 shows the sensitivity classification of selected model parameters based on five RSA runs of 5,000 simulations each. Key parameters are those classified as "critical" in at least three RSA runs. As expected of a lake eutrophication scenario, the key internal process parameters support the critical role of nutrient uptake and net primary production $\{\alpha_1, \alpha_2, \alpha_7, \alpha_8, \alpha_{49}, \alpha_{51}, \alpha_{52}, \alpha_{60}\}$, grazing (top-down) control $\{\alpha_4, \alpha_{13}, \alpha_{33}, \alpha_{36}\}$, and sediment-water-nutrient interactions $\{\alpha_{47}, \alpha_{56}\}$. Also, the initial conditions for phytoplankton and macrozooplankton $\{\alpha_{83}, \alpha_{86}\}$ indicate the importance of careful field measurements for state estimation. The pattern of stream inflow $\{\alpha_{101}\}$ – and by extension, the loading of nutrients, sediments and organic matter – allude to the critical influence of watershed hydrology and land use on water quality in lakes and reservoirs.

To prioritize management actions, or indeed direct the focus of further scientific research, it is essential to determine the relative importance of the key parameters that have been identified. The RSA provides a possible statistic for this. However, this task is better deferred to the TSDE because of its strong multivariate basis.

Table 2 presents the main attributes of the eight highest-density (of 18) terminal nodes in the TSDE tree diagram for RSA run #3. Identical results were obtained for the other four runs. The nodes contain a combined total of 84 (of 171) behavior-producing parameter sets, occupying just 6.3% of the RSA sampling domain – an indication of the probability of matching the specified behavior definition. Levels 1→8 trace the parameters that define each node. The sediment-water diffusion coefficient $\{\alpha_{56}\}$ is the most sensitive, since it defines the first split, suggesting that sediment-water-nutrient exchange is the most critical process. The zooplankton parameters $\{\alpha_{36}, \alpha_4, \alpha_{13}\}$ define all eight nodes, thus indicating that grazing is the next most critical process. Next, the initial phytoplankton biomass $\{\alpha_{83}\}$ defines five nodes, again an indication of the need for careful estimation from field measurements or prior models. The phosphorus half-saturation constant $\{\alpha_8\}$ supports nutrient uptake. More importantly, the inflow pattern $\{\alpha_{101}\}$ defines four nodes, and further supports the strong influence of watershed hydrology and land use. Lastly, the timing of insect emergence $\{\alpha_{115}\}$ defines two nodes, while epilimnion water temperature $\{\alpha_{103}\}$, detritus settling rate $\{\alpha_{50}\}$ and phytoplankton growth rate $\{\alpha_1\}$ each define one node.

Table 1. RSA sensitivity classification of selected process parameters, initial states and input/forcing functions [1 = critical 2 = important 3 = insignificant]

parameter	description	sensitivity class for run#: 1	2	3	4	5
Process parameters:						
α_1	phytoplankton growth rate constant	1	1	1	1	3
α_2	microbes growth rate constant	3	1	3	1	1
α_3	microzooplankton growth rate constant	2	3	3	3	1
α_4	macrozooplankton growth rate constant	1	1	1	1	1
α_7	half-sat. const.: phosphorus → phyt.	1	3	1	1	1
α_8	half-sat. const.: phosphorus → microbes	1	1	1	1	1
α_{11}	half-sat. const.: microbes → microzoop.	3	1	3	3	1
α_{13}	half-sat. const.: phyt. → macrozoop.	1	1	1	2	1
α_{33}	zooplankton carbon-biomass ratio	3	1	1	2	1
α_{36}	zooplankton loss rate constant	1	1	1	1	1
α_{47}	phosphorus desorption rate (sedimentlayer)	1	1	1	1	1
α_{49}	settling velocity of phytoplankton	1	1	1	3	1
α_{50}	settling velocity of detritus	3	3	3	3	3
α_{51}	temperature constant (growth)	1	3	1	3	1
α_{52}	temperature constant (respiration)	1	1	1	1	1
α_{56}	sediment-water diffusion coefficient	1	1	1	1	1
α_{60}	light extinction due to suspended solids	1	1	1	3	3
Initial states:						
α_{83}	initial condition: phytoplankton	1	1	1	1	1
α_{86}	initial condition: macrozooplankton	1	3	3	1	1
α_{88}	initial condition: larval-juvenile fish	3	3	3	3	3
Input/forcing functions:						
α_{101}	index (year) of inflow time series	1	1	2	1	1
α_{103}	mean water temperature (epilimnion)	3	3	3	3	3
α_{115}	time of onset of insect emergence	2	3	3	3	3

Table 2. High-density terminal nodes of the TSDE tree diagram

node	density	#points	volume [%]	tree level [1 = root]: 1	2	3	4	5	6	7	8
S_{31}	36.10	10	0.16	α_{56}	α_{36}	α_{83}	α_{13}	α_8	α_{101}		
S_{33}	28.08	7	0.15	α_{56}	α_{36}	α_{83}	α_{13}	α_8	α_{101}	α_{56}	
S_{34}	26.74	6	0.13	α_{56}	α_{36}	α_{83}	α_{13}	α_8	α_{101}	α_{56}	α_{115}
S_{25}	8.91	16	1.05	α_{56}	α_{36}	α_{83}	α_1	α_{50}			
S_{13}	8.35	10	0.70	α_{56}	α_{36}	α_4	α_{13}				
S_{35}	6.93	14	1.18	α_{56}	α_{36}	α_{83}	α_{13}	α_8	α_{101}	α_{56}	α_{115}
S_{14}	6.50	7	0.63	α_{56}	α_{36}	α_4	α_{13}	α_{88}			
S_{17}	3.61	14	2.27	α_{56}	α_{36}	α_4	α_{13}	α_{88}	α_{103}		
		84	**6.27**								

CONCLUSIONS

In this paper, we presented two sampling-based methods for uncertainty analysis – the Regionalized Sensitivity Analysis (RSA) and the Tree-Structured Density Estimation (TSDE) procedures. In a case study of Lake Lanier, we applied both methods to model-based predictions of lake eutrophication. Our results suggest that the major regulators of lake eutrophication include sediment-water-nutrient interactions, limited grazing, high nutrient loading and primary production.

While these results support well established theories, we have nonetheless demonstrated a logically simple (though computationally intensive) evaluation of the relative importance of major factors that influence the future behavior of a natural system. The RSA and TSDE procedures are potentially beneficial to planning and decision-making in natural resource management, most especially, in ranking priorities for policy actions.

ACKNOWLEDGMENTS

This study was funded by grant # R825758 under the EPA-NSF 1998 Water and Watersheds Program. Dr. Gerrit Hoogenboom of the Georgia Automated Environmental Monitoring Network (AEMN) provided the Gainesville solar radiation data.

LITERATURE CITED

Beck M. B. (1991). Principles of modelling. *Water Science & Technology*, 24(6): 1-8.

Hatcher, K.J., 1994 (ed.). *Project completion report: Diagnostic/feasibility study of Lake Lanier, Georgia.* Prepared for Georgia Environmental Protection Division, Atlanta, Georgia.

Osidele, O.O. and M.B. Beck, 1999. Identifying key model parameters in matching observed past and possible future behaviors for Lake Oglethorpe, Georgia. In: *Proceedings of the 1999 Georgia Water Resources Conference, held March 30-31,1999, at the University of Georgia*, K. J. Hatcher, editor, Institute Ecology, The University of Georgia, Athens, GA.

Simons, T.J. and D.C.L. Lam, 1980. Some limitations of water quality models for Large Lakes: A case study of Lake Ontario. *Water Resources Research*, 16(1): 105-116.

Spear, R.C., T.M. Grieb and N. Shang, 1994. Parameter uncertainty and interaction in complex environmental models. *Water Resources Research*, 30(11):3159-3169.

Spear, R.C. and G.M. Hornberger, 1980. Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, 14:43-49.